Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2017 vol.10089 LNAI, pages 174-184

Automated prediction of demographic information from medical user reviews

Tutubalina E., Nikolenko S.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2017, Springer International Publishing AG.The advent of personalized medicine and wide-scale drug tests has led to the development of methods intended to automatically mine and extract information regarding drug reactions from user reviews. For medical purposes, it is often important to know demographic information on the authors of these reviews; however, existing studies usually either presuppose that this information is available or disregard the issue. We study automatic mining of demographic information from user-generated texts, comparing modern natural language processing techniques, including extensions of topic models and deep neural networks, for this problem on datasets mined from health-related web sites.

http://dx.doi.org/10.1007/978-3-319-58130-9_17

References

[1] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. 3(4–5), 993–1022 (2003)

[2] Feldman, R., Netzer, O., Peretz, A., Rosenfeld, B.: Utilizing text mining on online medical forums to predict label change due to adverse drug reactions. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2015, New York, pp. 1779–1788. ACM (2015)

[3] Goldberg, Y.: A primer on neural network models for natural language processing. CoRR abs/1510.00726 (2015). http://arxiv.org/abs/1510.00726

[4] Griffiths, T., Steyvers, M.: Finding scientific topics. Proc. Natl. Acad. Sci. 101(Suppl.1), 5228–5335 (2004)

[5] Karimi, S., Wang, C., Metke-Jimenez, A., Gaire, R., Paris, C.: Text and data mining techniques in adverse drug reaction detection. ACM Comput. Surv. 47(4), 56:1–56:39 (2015)

[6] Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint (2014). arXiv:1408.5882

[7] Leaman, R., Wojtulewicz, L., Sullivan, R., Skariah, A., Yang, J., Gonzalez, G.: Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In: Proceedings 2010 Workshop on Biomedical Natural Language Processing, BioNLP 2010, pp. 117–125. ACL, USA (2010)

[8] Marcus, A.D.: Researchers fret as social media lift veil on drug trials. Wall Street J. (2014). http://www.wsj.com/articles/researchers-fret-as-social-media-lift-veil-on-drug-trials-1406687404

[9] Martínez, P., Martínez, J.L., Segura-Bedmar, I., Moreno-Schneider, J., Luna, A., Revert, R.: Turning user generated health-related content into actionable knowledge through text analytics services. Comput. Ind. 78, 43–56 (2016)

[10] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. CoRR abs/1301.3781 (2013). http://arxiv.org/abs/1301.3781

[11] Mikolov, T., Karafiá, M., Burget, L., Cernocký, J., Khudanpur, S.: Recurrent neural network based language model. Interspeech 2, 3 (2010)

[12] Mikolov, T., Kombrink, S., Burget, L., Vcernocký, J.H., Khudanpur, S.: Extensions of recurrent neural network language model. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5528–5531. IEEE (2011)

[13] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. CoRR abs/1310.4546 (2013). http://arxiv.org/abs/1310.4546

[14] Mnih, A., Hinton, G.E.: A scalable hierarchical distributed language model. In: Advances in neural information processing systems, pp. 1081–1088 (2009)

[15] Plachouras, V., Leidner, J.L., Garrow, A.G.: Quantifying self-reported adverse drug events on twitter: signal and topic analysis. In: Proceedings of the 7th 2016 International Conference on Social Media & Society, SMSociety 2016, New York, pp. 6:1–6:10. ACM (2016)

[16] Ramage, D., Manning, C.D., Dumais, S.: Partially labeled topic models for interpretable text mining. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 457–465. ACM (2011)

[17] Rastegar-Mojarad, M., Liu, H., Nambisan, P.: Using social media data to identify potential candidates for drug repurposing: a feasibility study. JMIR Res. Protoc. 5(2), e121 (2016)

[18] Sarker, A., Nikfarjam, A., Gonzalez, G.: Social media mining shared task workshop. In: Proceedings of Pacific Symposium on Biocomputing 2016, pp. 581–592 (2016)

[19] Segura-Bedmar, I., Martínez, P., Revert, R., Moreno-Schneider, J.: Exploring spanish health social media for detecting drug effects. BMC Med. Inf. Decis. Making 15(2), 1–9 (2015)

[20] Shaywitz, D., Mammen, M.: The next killer app. The Boston Globe (2011). http://archive.boston.com/bostonglobe/editorial opinion/oped/articles/2011/01/23/thenext killer app/

[21] Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phraselevel sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354. Association for Computational Linguistics (2005)

[22] Yang, C.C., Yang, H., Jiang, L., Zhang, M.: Social media mining for drug safety signal detection. In: Proceedings of the 2012 International Workshop on Smart Health and Wellbeing, SHB 2012. NY, USA, pp. 33–40 (2012). http://doi.acm.org/10.1145/2389707.2389714

[23] Z. Yang, A. Kotov, A.M., Lu, S.: Parametric and non-parametric user-aware sentiment topic models. In: Proceedings of the 38th ACM SIGIR (2015)

[24] Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, NIPS 2015, pp. 649–657. MIT Press, Cambridge (2015). http://dl.acm.org/citation.cfm?id=2969239.2969312

[25] Zhang, Z., Nie, J.Y., Zhang, X.: An ensemble method for binary classification of adverse drug reactions from social media. In: Proceedings of the Social Media Mining Shared Task Workshop at the Pacific Symposium on Biocomputing (2016)